DOCUMENT RESUME

ED 259 000 TM 850 369

AUTHOR Bradley, Drake R.
TITLE Computer Simulation of Classic Studies in Psychology.
PUB DATE Apr 84
NOTE 22p.; Paper presented at the Annual Meeting of the Eastern Psychological Association (Baltimore, MD, April 12-15, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Computer Oriented Programs; *Computer Simulation; *Courseware; Higher Education; *Monte Carlo Methods; Programed Instructional Materials; *Psychological Studies; Reliability; *Research Design; Statistical Data
IDENTIFIERS *DATASIM (Computer Software)

ABSTRACT

This paper describes DATASIM, a comprehensive software package which generates simulated data for actual or hypothetical research designs. DATASIM is primarily intended for use in statistics and research methods courses, where it is used to generate "individualized" datasets for students to analyze, and later to correct their answers. The generality of its design allows DATASIM to be used to conduct Monte Carlo studies. The methods employed generate datasets whose statistics follow the appropriate sampling distributions. This is true provided that the distribution shape of the variable to be simulated is normal, and that the investigator has not set scale limits on the generation of the data in order to avoid "off range" values. The generation of simulated data by DATASIM for a particular research design is determined by parameters stored in an external file on disk. The application of simulation methods to assess power is examined for several classic studies. The appendices present results of several simulated datasets of the Aronson and Mills (1959) study of the effects of severity of initiation on liking for a group and the Miller and DiCara (1967) study on the instrumental conditioning of heart rate response in rats. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED259000

SHORT ABSTRACT

1) Title: Computer Simulation of Classic Studies in Psychology.

2) Author: Drake R. Bradley, Bates College

3) Sponsor: none

4) Text: A comprehensive software package, called DATASIM, allows investigators to generate, display, and analyze simulated data for a wide variety of research designs, from simple two group experiments to repeated measures factorial designs, multivariate research, and research producing counted data (contingency tables). The application of simulation methods to assess power is examined for several classic studies.

5) Topical Session Preference: Quantitative-Methodological

6) Slides and/or transparencies will be used.

7) Drake R. Bradley
Department of Psychology
Bates College
Lewiston, Maine, 04240

Please reference this paper as follows:

Bradley, D. R. Computer simulation of classic studies in psychology. Proceedings and Abstracts of the Eastern Psychological Association, 1984, 55, 60.

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Bradley, Drake R.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 850 369

LONG ABSTRACT

1) Title: Computer Simulation of Classic Studies in Psychology.

2) Problem: A number of the traditional findings in psychology are based on research designs employing small samples. If the effects under investigation are relatively weak, the power of the statistical test will be low, and subsequent investigators may find it difficult to replicate the original study. Rossi (1982, 1983) has suggested that a number of artifactual controversies -- for example, that over spontaneous recovery -- have resulted from this state of affairs. Consequenty, the stability of a particular pattern of results reported in the literature becomes an important factor in the interpretation of conflicting findings by subsequent investigators.

An extremely useful approach for gaining insight into the problem of "stability" of research results relies on computer simulation or modeling. Assume for the moment that a computer program exists which can generate simulated data for most research designs employed by psychologists, and that the results may be conveniently analyzed and displayed. Furthermore, assume that to use the program an investigator need only specify relevant parameters for the particular application involved. For example, to simulate the classic study by Aronson and Mills (1959) -- on the effect of severity of initiation on liking for a group -- the parameters to be specified are the population means and standard deviations, and the sample sizes, for the three conditions of the experiment. The program would then generate successive sets of simulated data, plot the results, and perform any statistical analyses requested by the user. Visual inspection of the plots, and examination of the p-levels of the F-tests performed, would then give the user an excellent indication of the inherent stability of the pattern of

results originally reported by Aronson and Mills (1959).

Consider the following related problem. Suppose an investigator predicts that an experiment will have a particular pattern of results, and that a two-way ANOVA on the data will have certain effects being significant and other effects being nonsignificant (of the possible main, interaction, and simple effects which could be tested). The investigator is interested in whether the sample size to be used provides adequate power for detecting these various effects, and more generally in to extent to which the predicted pattern will replicate in subsequent investigations. While actual power calculations can be made for the separate effects to be tested, it should be evident that having the computer simulate "doing the experiment" over and over again provides similar information in a form that is more accessible. The investigator can sit back and view the plotted results and the F-tests for each simulated dataset. The sample size could be altered, and the simulation conducted again, until the investigator has a "seat-of-the-pants" feel for the stability of the results.

The present paper describes a comprehensive software package which generates simulated data for actual or hypothetical research designs. The package, called DATASIM, is primarily intended for use in statistics and research methods courses, where it is used to generate "individualized" datasets for students to analyze, and later to correct their answers. However, the generality of its design allows DATASIM to be used to conduct Monte Carlo studies, and to answer questions like those posed above. The remaining sections of this paper are devoted to illustrating its use in the latter context.

3) Method: A discussion of the computational algorithms used by DATASIM to generate univariate and multivariate datasets, as well as

datasets for categorical variables (contingency tables, etc), is beyond the scope of the present paper. The methods employed are similar to those discussed by Lehman (1977, Chap. 5), and generate datasets whose statistics (means, standard deviations, correlations, regression coefficients, F, t, etc.) follow the appropriate sampling distributions. This is true provided that the distribution shape of the variable to be simulated is normal, and that the investigator has not set scale limits on the generation of the data in order to avoid "off-range" values. (While the current version of DATASIM, written in BASIC, only generates normally distributed data, a new version, written in PL1G, uses the lambda distribution (Ramberg & Schmeiser, 1974) to generate data from unimodal symmetric and nonsymmetric distributions having whatever degree of skewness and kurtosis is desired by the user.)

The generation of simulated data by DATASIM for a particular research design is determined by parameters stored in an external file on disk. This "driver" file contains other information as well, used for labeling conditions, controlling plots of the data, and so on. Consider the following driver file which controls the simulations for the Aronson and Mills (1959) experiment discussed above:

```
3   [number of groups or variables]
LABELS,CONTROL,MILD INITIATION,SEVERE INITIATION,* [cond. labels]
MU,170.1,171.1,195.3,*   [population means]
SIGMA,21.6,34,31.9,*   [population s.d.'s]
SIZE,21,*   [sample size]
XYLABELS,LIKING,INITIATION CONDITION,*   [plotting labels]
PLOT MEAN,150,10,7,0,1,4,1,*   [plotting info]
PLOT RAWDATA,50,50,4,0,1,4,.9,*   [plotting info]
REFERENCE, ARONSON & MILLS (1959), JASP, *   [reference]
```

The only information essential to the simulation in this driver file is that specified in the first line (K, the number of conditions)

and in the lines labeled MU (population means), SIGMA (population S.D.s), and SIZE (sample size per condition). The remaining information is used to control labeling and plotting, and can be omitted (default values are the used).

It is also possible to specify upper and/or lower limits for the data in each condition of a study, so as to avoid generating off-range values. These should be omitted when accurate simulation is desired because the corresponding sampling distributions are affected when off-range values are replaced by on-range values. Of course, the existence of this problem implies that the original data (for the study being simulated) was not normally distributed: i.e., the mean of a sample was less than three standard deviations away from the upper (or lower) scale limit.

Multivariate datasets and datasets for experimental designs having repeated measures require that the driver file also specify the population correlation matrix. Datasets for categorical data (one or twoway classification) require that population probabilities be specified for the various categories or cells. Regardless of the specific research design involved, the investigator will usually need to use the sample statistics of a previous study to estimate these values. In the example shown above, the "population" means and standard deviations were obtained from Aronson and Mills (1959). Since many published studies do not provide the needed information directly, it is often necessary to determine these values on the basis of information that is provided (e.g., as in using the mean square error in an ANOVA to estimate individual cell variances). And in some cases, one must simply make an intelligent guess as to a required parameter (e.g., the correlation between levels of a repeated measures factor).

In any event, to simulate a classic (or any other) study in psychology, the user creates a text file containing estimates of the appropriate parameters, and then executes DATASIM and supplies it with the name of that file. DATASIM reads the information from the driver file, initializes values in the program, and then awaits commands from the user dictating the number of simulations to be conducted, and the plots and/or analyses to be conducted for each simulated dataset.

4) Results: Appendix 1 present the results of several simulated datasets of the Aronson and Mills (1959) study. An inspection of the plots shows that the pattern of results is reasonably stable across replications, and the F-test conducted is usually significant. Hence, exact replications of this study using samples of 21 subjects in each condition should have a good chance of supporting the original results. An empirical estimate of the power of the F-test is given by the proportion of datasets resulting in significant F-tests. In the present instance, 80 of the 100 simulated datasets (not all of which are shown) resulted in significant F-tests, so the power of the F-test would seem adequate.

In contrast to the Aronson and Mills (1959) study, one conducted by Festinger and Carlsmith (1959) has a pattern of results which is considerably less stable from one replication to the next. Out of 100 simulated datasets, only 57 resulted in significant t-tests when comparing the $1 condition to the $20 condition. Since this empirical estimate of power (.57) is close the .50 value discussed by Rossi (1982, 1983), there is good reason to expect "artifactual" controversy will surround this particular finding (and indeed, to some considerable extent, this has been the case with this and other dissonance studies).

As one last example, we consider a simulation of the classic study

by Miller and DiCara (1967), which demonstrated the operant conditioning of heart rates in rats. The design was a 2 x 4 factorial design with repeated measures on one factor. One group of rats was rewarded for increasing their heart rates, while a second group was rewarded for decreasing their heart rates. Both groups were assessed after 0, 30, 60, or 90 minutes of training. Appendix 2 presents the results of a number of simulated datasets. In order to conduct these simulations, it was necessary to estimate the correlation between levels of the training factor (r = .8 was used). Also the variance within groups had to be estimated from t-tests reported by the authors. In any event, it should be apparent that inspection of the plots and the twoway ANOVAs conducted provide us with a reasonably good feel for the stability of the results.

5) Discussion: While power calculations could provide us with the same information as that conveyed by computer simulation, the latter method is not only more convenient but has greater flexibility and generality. For example, techniques for computing power assume a common population variance and frequently assume equal sample sizes as well. If an investigator wants to replicate a previous study in which there is good evidence that the variances are not equal, or in which unequal sample sizes have been employed, this need not present any obstacles to estimating power if computer simulation is employed. The (unequal) sample standard deviations and/or sample sizes are simply specified in the driver file (as obtained from the original study), the remaining information provided, and the simulation conducted. Also, using computer simulation to estimate power is no more difficult for complex experimental designs (e.g., mixed model ANOVA) than for simple two group designs, since in either case the simulation is controlled by

a text file containing only a few lines of information. Calculating power, on the other hand, becomes increasingly more involved for complex experimental designs, especially if one is interested in the power of individual post hoc commparisons as well as that of the overall F-tests. Finally, it quite easy to change certain parameters in the driver file (e.g., the sample sizes, the correlations between groups, the magnitude of the experimental effects, etc.) and rerun the simulation, whereas recomputing exact power values for each change in parameter(s) would be extremely tedious.

References

Aronson, E. & Mills, J. Effect of severity of initiation on liking for a group. The Journal of Abnormal and Social Psychology, 1959, 59, 177-181.

Festinger, L. & Carlsmith, J. M. Cognitive consequences of forced compliance. Journal of Abnormal and Social Psychology, 1959, 58, 203-210.

Lehman, R. S. Computer simulation and modeling. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1977.

Miller, N. E. & DiCara, L. Instrumental learning of heart rate changes in curarized rats: shaping, and specificity to discriminative stimulus. Journal of Comparative and Physioloical Psychology, 1967, 63, 12-19.

Ramberg, J. S., & Schmeiser, B.W. An approximate method for generating asymmetric random variables. Management Science / Operations Research (ACM), 1974, 17, 78-82.

Rossi, J.S. Meta-analysis, power analysis and artifactual controversy: The case of spontaneous recovery of verbal associations. Eastern Psychological Association, Baltimore, April 1982.

Rossi, J. S. Inadequate statistical power: A source of artifactual controversy in behavioral research. Eastern Psychological Association, Philadelphia, April 1983.

```
DATASIM REVISION 09.29.83
USER N01057 ON 110483 AT 1243

? describe initiate
```

Aronson and Mills (1959) investigated the effect of severity of initiation on liking for a group. A sample of 63 college women volunteered to participate in small groups discussing the psychology of sex. Each woman who wished to become a member of a group listened to a tape recording of the group discussing the sexual behavior of lower animals. The discussion was deliberately designed to be as boring and banal as possible. The woman then rated the group discussion and the group members on 17 different bipolar scales (dull-interesting, unintelligent-intelligent, etc.) each of which ranged from 0 to 15 (with 15 being the positive end of the scale).

Prior to listening to the group discussion, each woman was required to take an "embarassment test." In the CONTROL condition, the women were simply asked whether or not they felt they could discuss sex freely and openly in the group. In the MILD INITIATION condition, the women were required to read aloud five words which were related to sex but which were not obscene. In the SEVERE INITIATION condition, the women were required to read aloud 12 obscene words, and two vivid descriptions of sexual activity from contemporary novels. Following the "initiation" test, the women listened to the group discussion and made their ratings. The ratings for the 17 scales were then summed, yielding total values which could range between 0 and 255.

REFERENCE

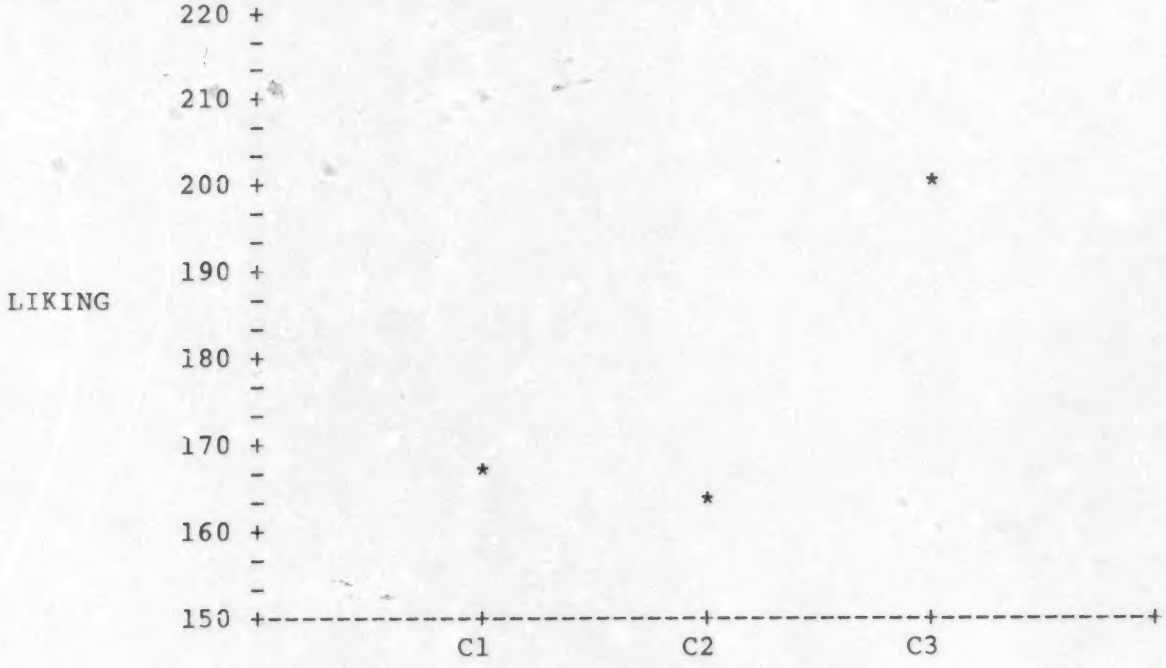
Aronson, E. & Mills, J. Effect of severity of initiation on liking for a group. The Journal of Abnormal and Social Psychology, 1959, 59, 177-181.

```
? simulate 10, plot means, oneway

SEED = 2251

          220 +
              -
              -
          210 +
              -
              -
          200 +                                          *
              -
              -
          190 +
LIKING        -
              -
          180 +
              -                      *
              -          *
          170 +
              -
              -
          160 +
              -
              -
          150 +----------+----------+----------+----------+
                         C1         C2         C3
```

INITIATION CONDITION

F-RATIO, NUMERATOR D.F., DENOMINATOR D.F., AND PROBABILITY:
==> 4.456, 2, 60, .0155

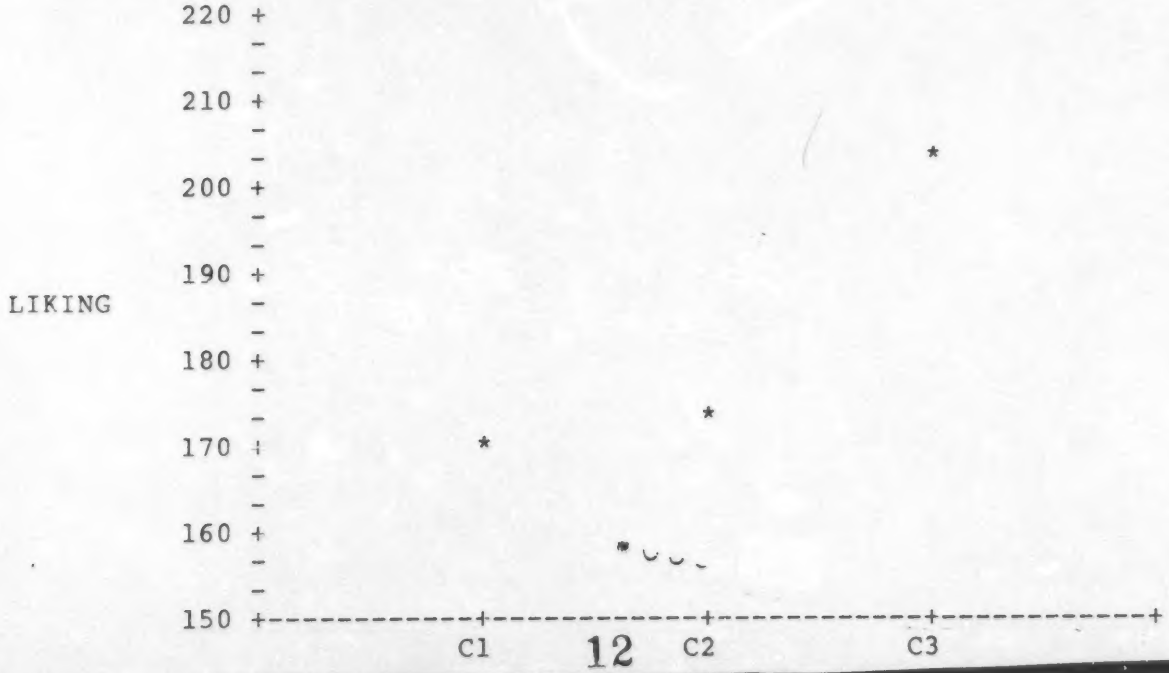
SEED = 2427



INITIATION CONDITION

F-RATIO, NUMERATOR D.F., DENOMINATOR D.F., AND PROBABILITY:
==> 12.145, 2, 60, 5.32E-05

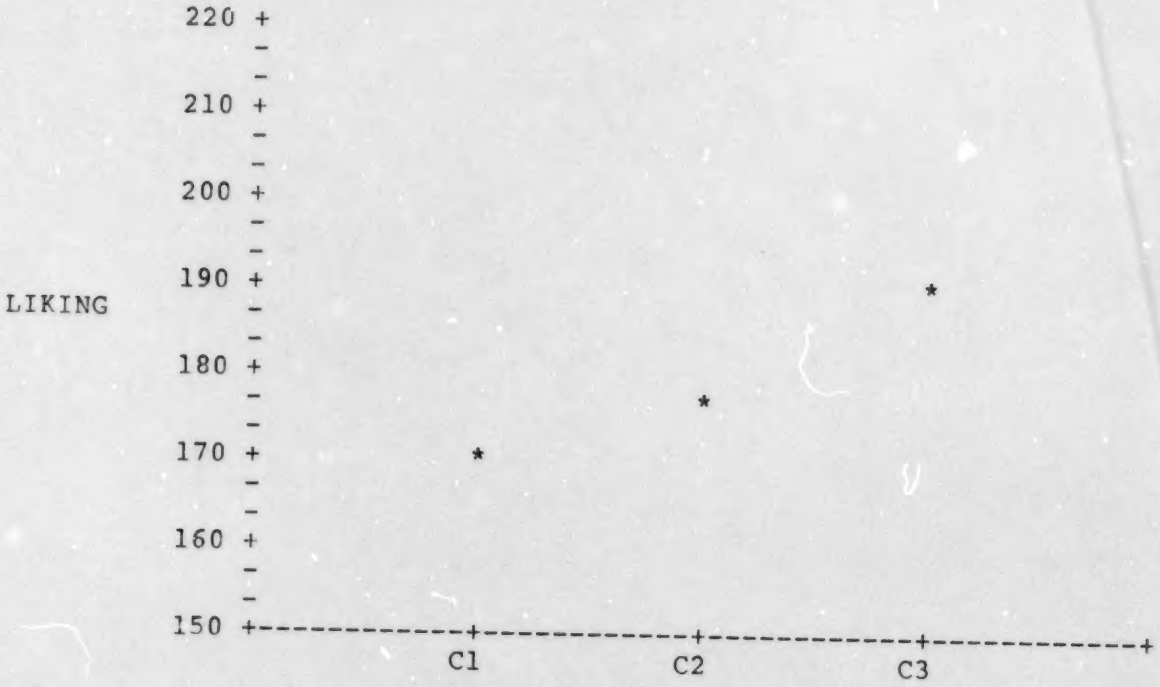
SEED = 2867



INITIATION CONDITION

F-RATIO, NUMERATOR D.F., DENOMINATOR D.F., AND PROBABILITY:
==> 6.679, 2, 60, .00255

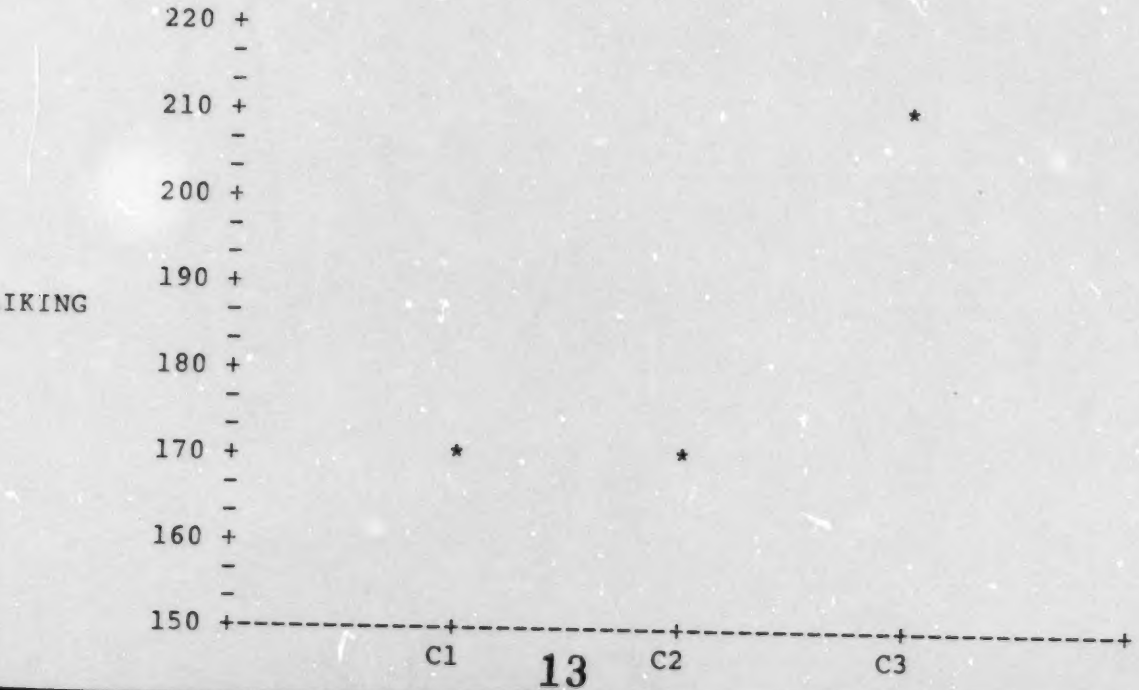
SEED = 8473



INITIATION CONDITION

F-RATIO, NUMERATOR D.F., DENOMINATOR D.F., AND PROBABILITY:
==> 2.756, 2, 60, .07

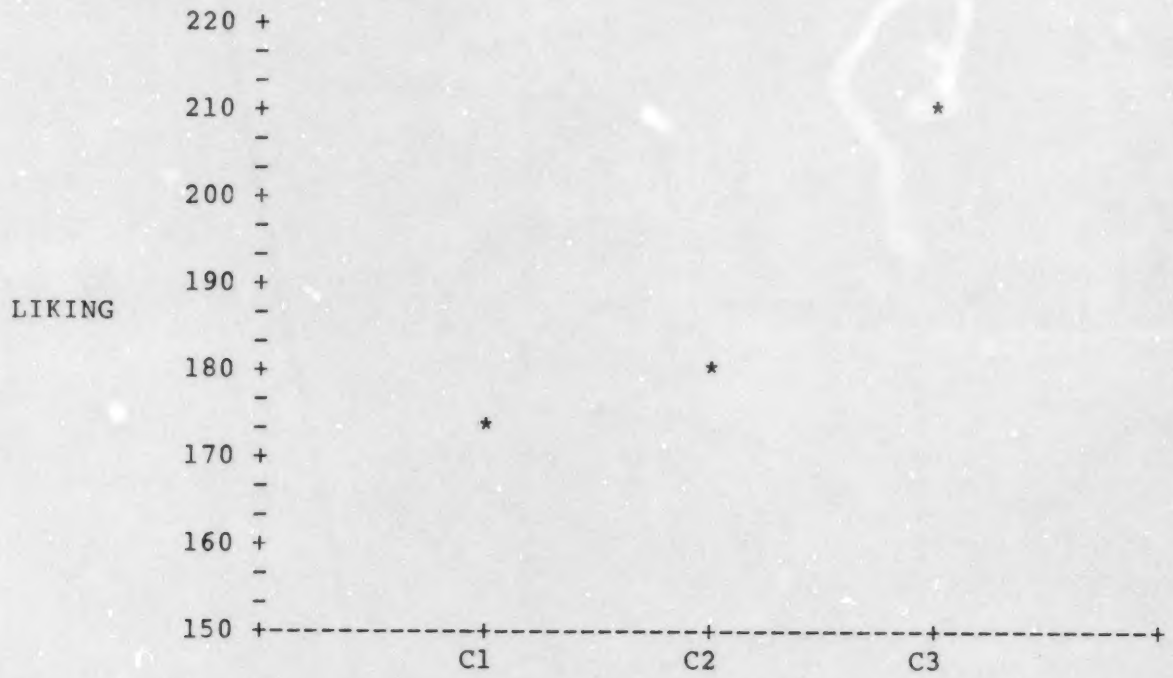
SEED = 5327



INITIATION CONDITION

F-RATIO, NUMERATOR D.F., DENOMINATOR D.F., AND PROBABILITY:
==> 14.777, 2, 60, 1.03E-05

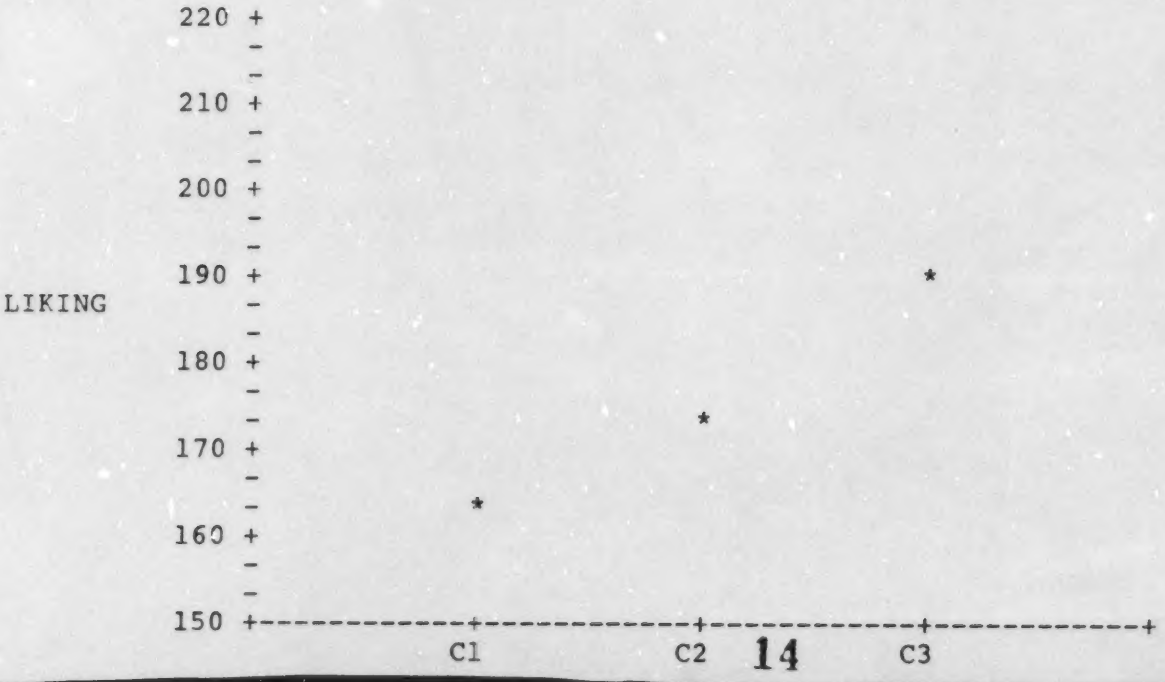
SEED = 9019



INITIATION CONDITION

F-RATIO, NUMERATOR D.F., DENOMINATOR D.F., AND PROBABILITY:
==> 7.103, 2, 60, .00184

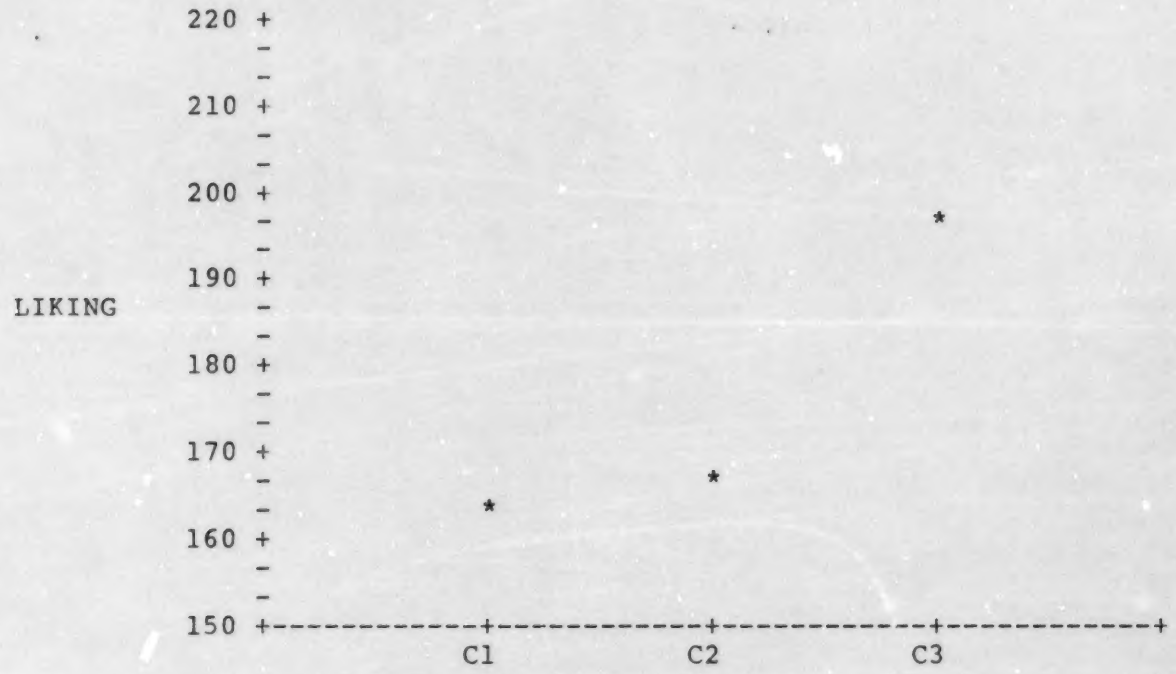
SEED = 5195



INITIATION CONDITION

F-RATIO, NUMERATOR D.F., DENOMINATOR D.F., AND PROBABILITY:
==> 5.145, 2, 60, .0087

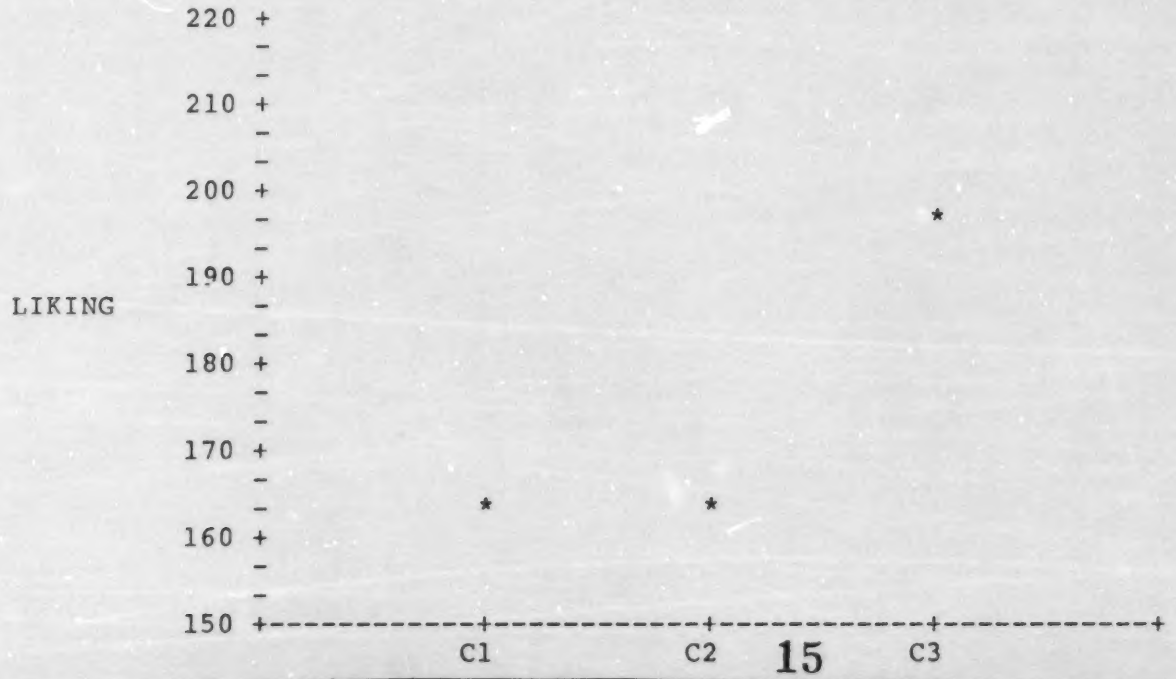
SEED = 9845



INITIATION CONDITION

F-RATIO, NUMERATOR D.F., DENOMINATOR D.F., AND PROBABILITY:
==> 5.723, 2, 60, .00543

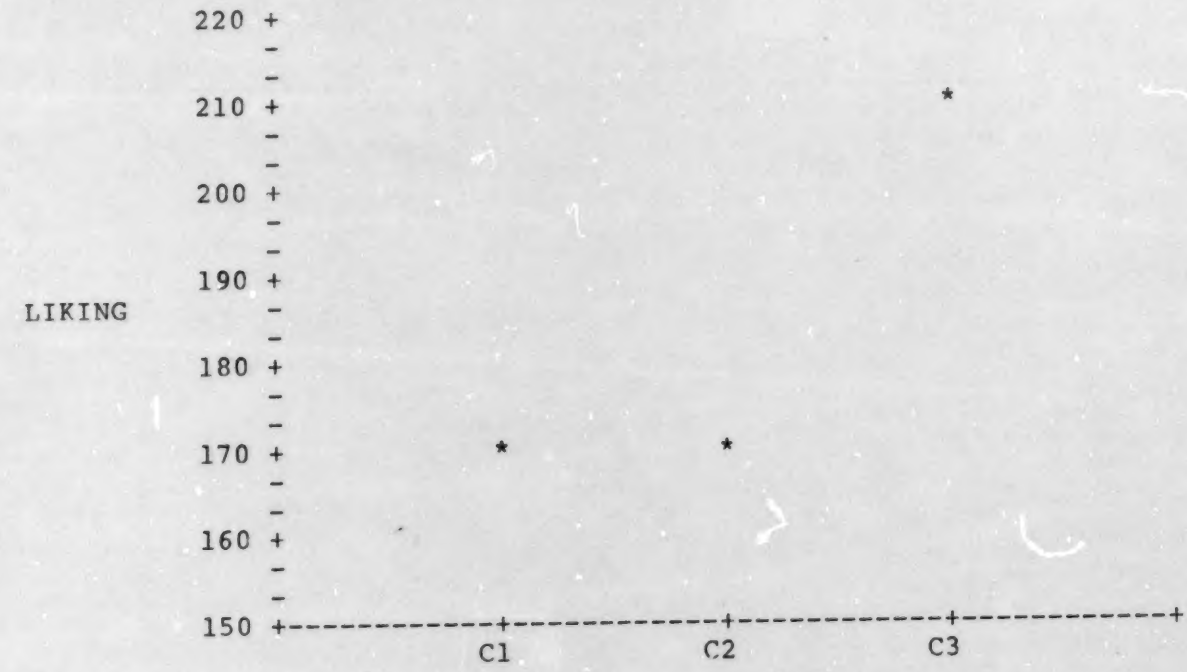
SEED = 3183



INITIATION CONDITION

F-RATIO, NUMERATOR D.F., DENOMINATOR D.F., AND PROBABILITY:
==> 7.647, 2, 60, .00122

SEED = 5327



INITIATION CONDITION

F-RATIO, NUMERATOR D.F., DENOMINATOR D.F., AND PROBABILITY:
==> 14.777, 2, 60, 1.03E-05

APPENDIX 2

? describe heart

In 1967 Miller and DiCara published a now classic paper on the instrumental (operant) conditioning of heart rate responses in rats. It had been previously supposed that "visceral" responses (e.g. heart rate, blood pressure, GSR, etc) were only susceptible to classical (Pavlovian) conditioning. Skeletal (motor) responses were susceptible to both classical and instrumental conditioning.

Miller and DiCara surgically implanted micro-electrodes into the medial forebrain bundle of each rat, and established that electrical stimulation of this region could result in the acquisition of an instrumental response (a bar press in this case). Each rat was then subjected to heart rate conditioning (see paper for actual procedure used). One group of rats (n = 12) was rewarded for increasing the heart rate and another (n = 12) for decreasing the heart rate. The rats heart rates were measured after 0, 30, 60, and 90 minutes of training. Hence, the design of the experiment was a 2 x 4 factorial design with a total of 8 different experimental conditions. The independent variables were the heart rate condition (increasing vs decreasing) and the amount of training provided (0, 30, 60, 90 minutes), and the dependent variable was the heart rate of the rat in beats per minute (BPM).

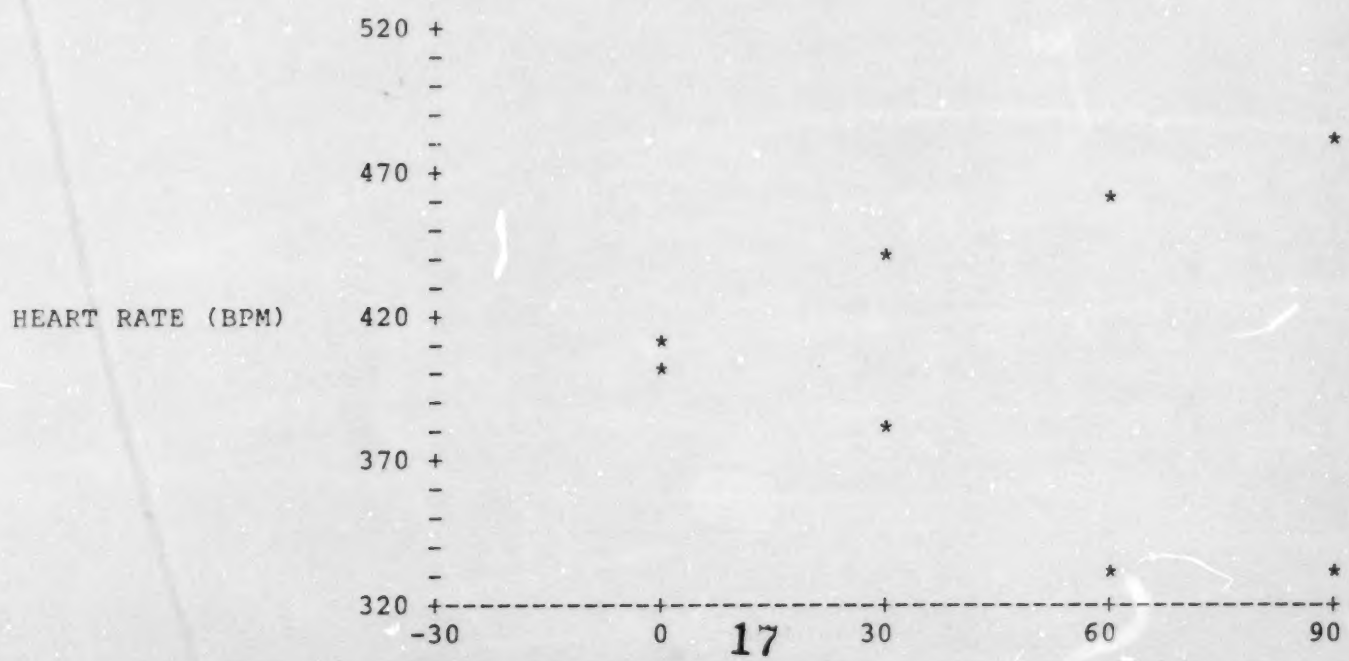
Have DATASIM generate simulated data for this experiment, and analyze the data in whatever ways you think are appropriate. Have your answers checked by DATASIM, and have a plot of the means made. Do you see any evidence for instrumental conditioning of heart rate in the data?

REFERENCE

Miller, N. E. & DiCara, L. Instrumental learning of heart rate changes in curarized rats: shaping, and specificity to discriminative stimulus. Journal of Comparative and Physioloical Psychology, 1967, 63, 12-19.

? simulate 10, plot means, twoway

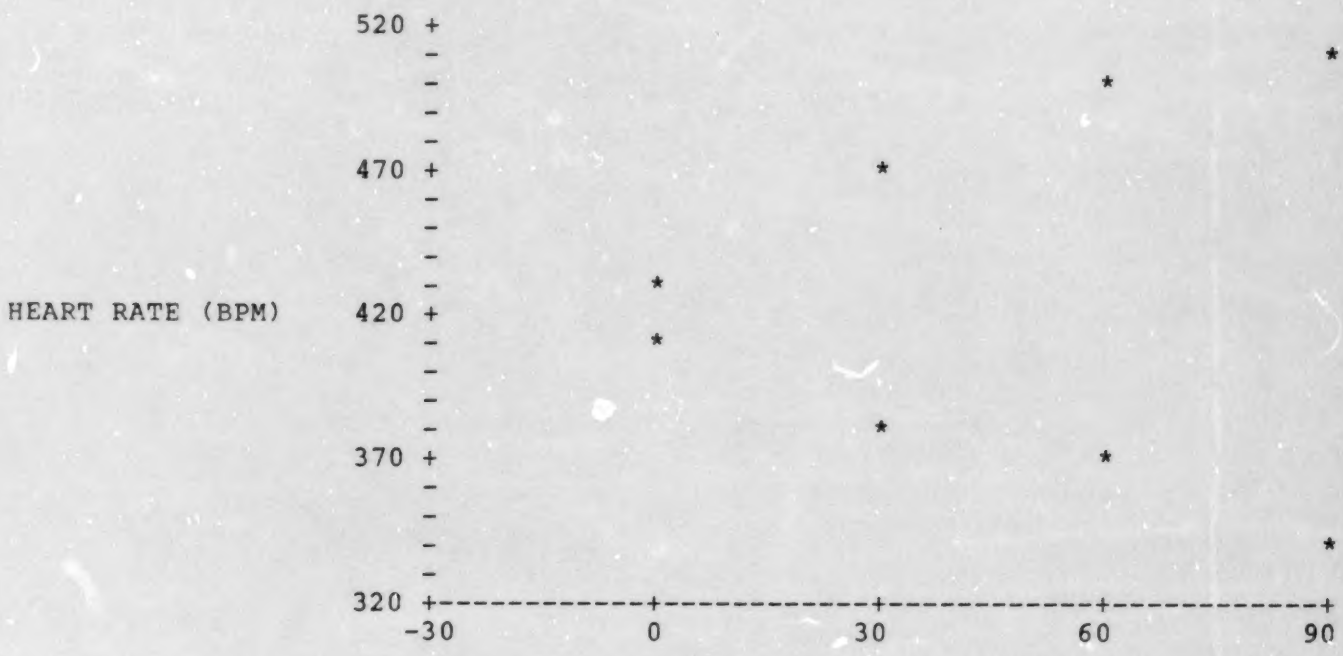
SEED = 8157



MINUTES OF TRAINING

F-RATIO, NUMERATOR DF, DENOMINATOR DF, AND PROBABILITY FOR:
REWARD (FAST/SLOW) ==> 4.24, 1, 22, .0515
TRAINING (0-90 MIN) ==> .424, 3, 66, .74
INTERACTION ==> 17.478, 3, 66, 5.11E-08

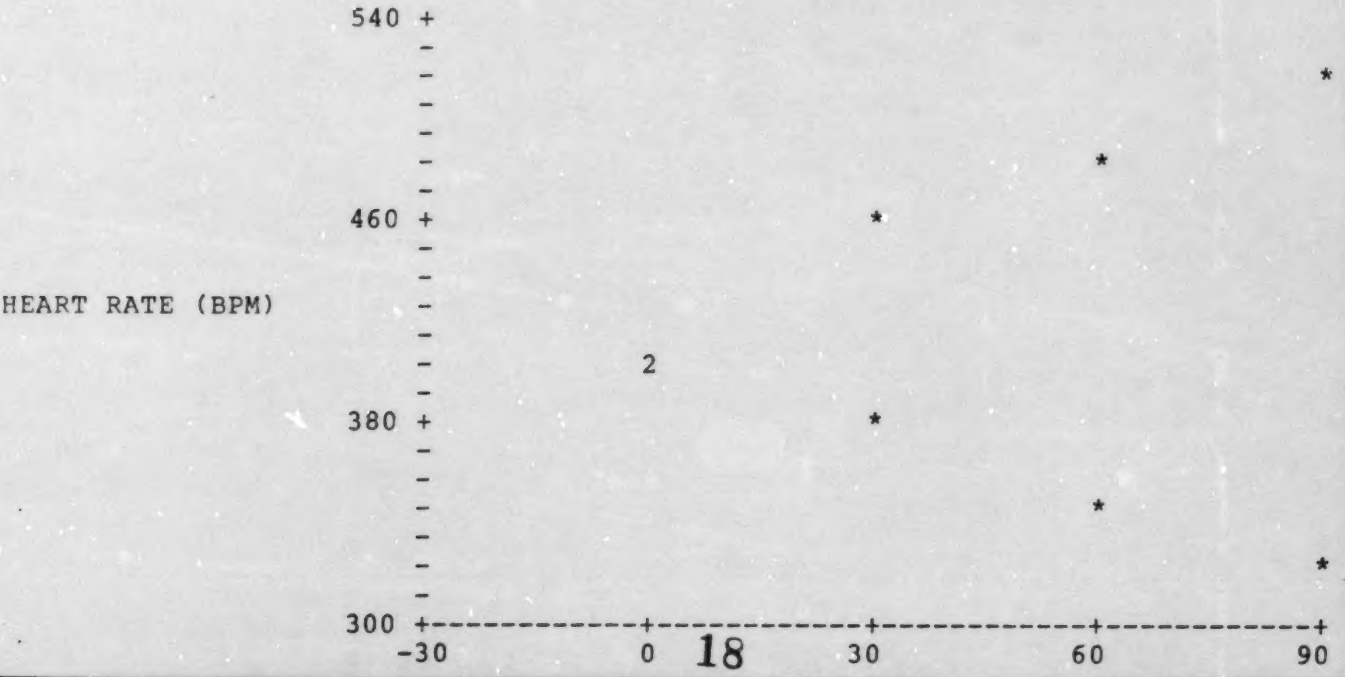
SEED = 1301



MINUTES OF TRAINING

F-RATIO, NUMERATOR DF, DENOMINATOR DF, AND PROBABILITY FOR:
REWARD (FAST/SLOW) ==> 15.215, 1, 22, .000768
TRAINING (0-90 MIN) ==> .61, 3, 66, .615
INTERACTION ==> 9.827, 3, 66, 2.68E-05

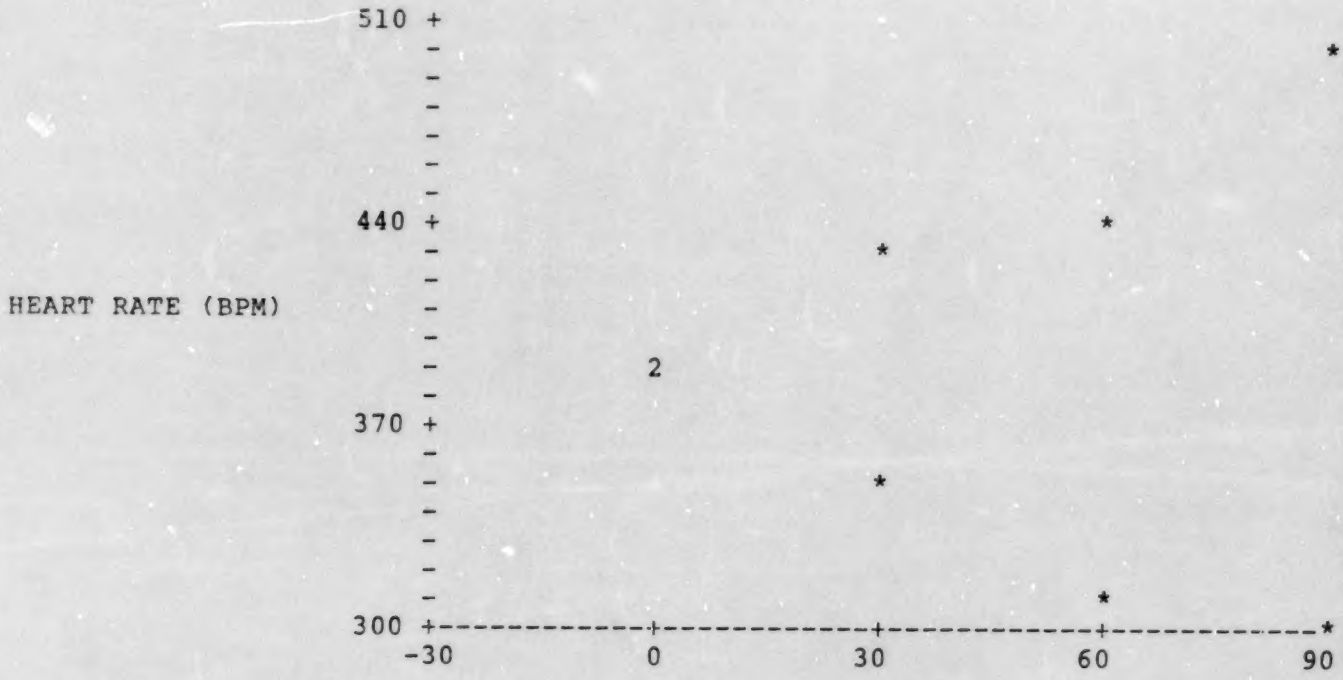
SEED = 5837



MINUTES OF TRAINING

F-RATIO, NUMERATOR DF, DENOMINATOR DF, AND PROBABILITY FOR:
REWARD (FAST/SLOW) ==> 7.791, 1, 22, .0106
TRAINING (0-90 MIN) ==> 1.027, 3, 66, .387
INTERACTION ==> 21.42, 3, 66, 3.46E-09

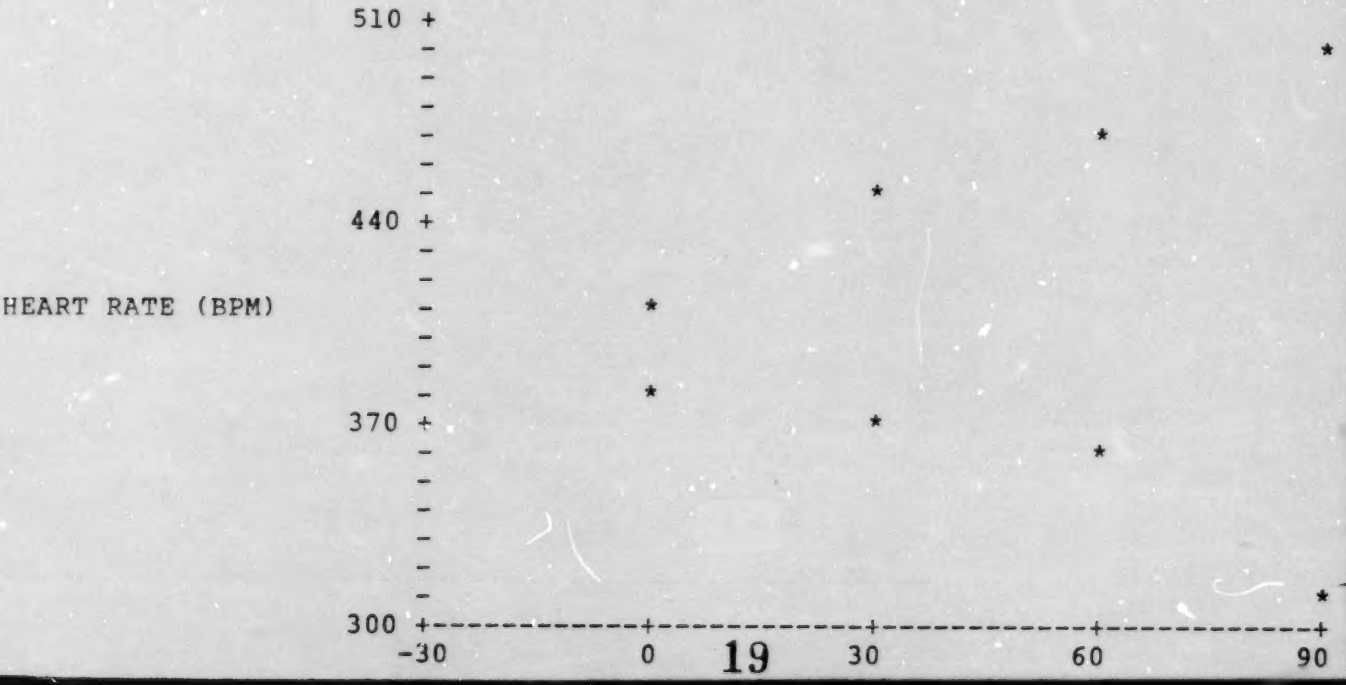
SEED = 3113



MINUTES OF TRAINING

F-RATIO, NUMERATOR DF, DENOMINATOR DF, AND PROBABILITY FOR:
REWARD (FAST/SLOW) ==> 9.842, 1, 22, .00479
TRAINING (0-90 MIN) ==> .897, 3, 66, .45
INTERACTION ==> 16.189, 3, 66, 1.32E-07

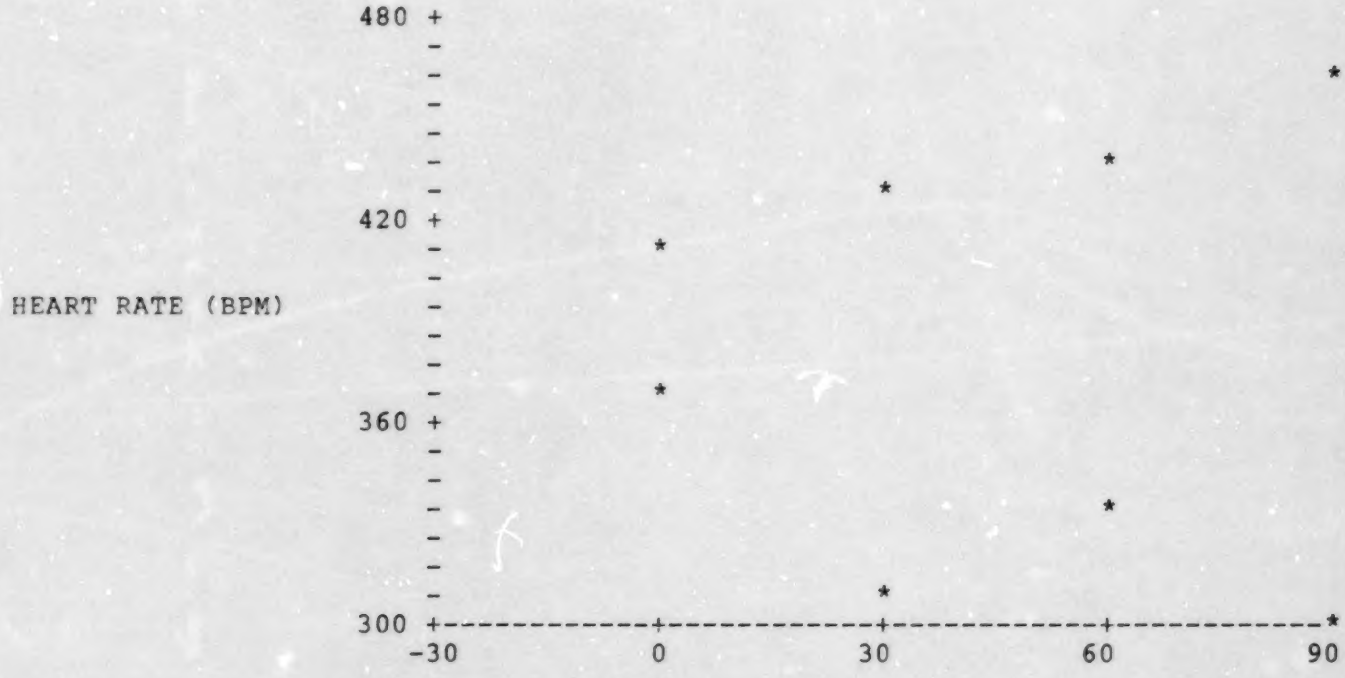
SEED = 7833



MINUTES OF TRAINING

F-RATIO, NUMERATOR DF, DENOMINATOR DF, AND PROBABILITY FOR:
REWARD (FAST/SLOW) ==> 9.839, 1, 22, .00479
TRAINING (0-90 MIN) ==> 1.185, 3, 66, .322
INTERACTION ==> 14.982, 3, 66, 3.33E-07

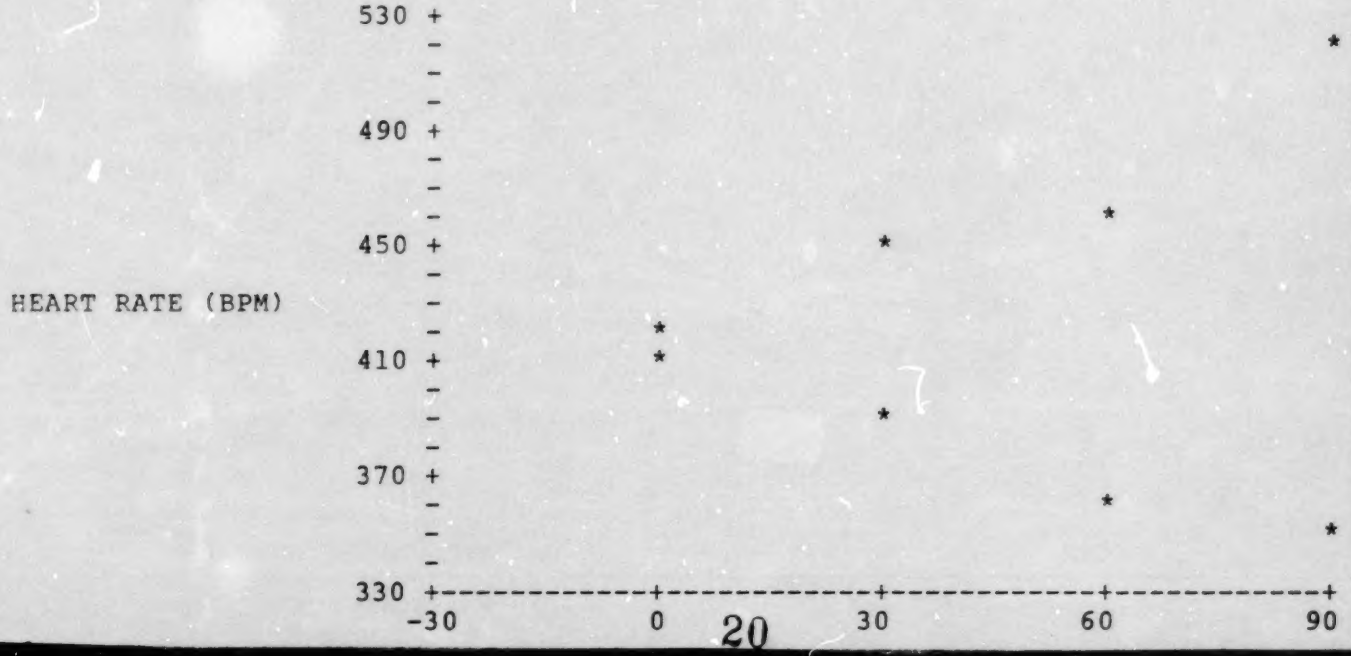
SEED = 4141



MINUTES OF TRAINING

F-RATIO, NUMERATOR DF, DENOMINATOR DF, AND PROBABILITY FOR:
REWARD (FAST/SLOW) ==> 5.722, 1, 22, .0257
TRAINING (0-90 MIN) ==> 1.249, 3, 66, .299
INTERACTION ==> 9.232, 3, 66, 4.69E-05

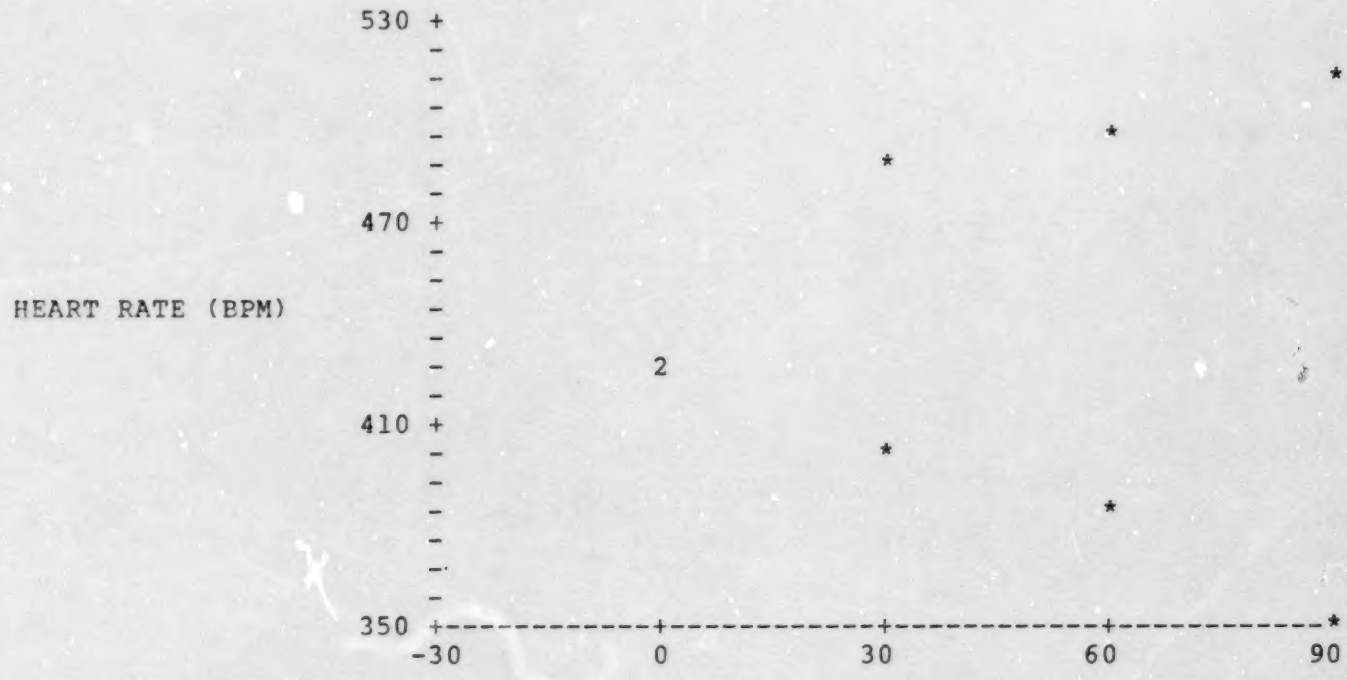
SEED = 5503



MINUTES OF TRAINING

F-RATIO, NUMERATOR DF, DENOMINATOR DF, AND PROBABILITY FOR:
REWARD (FAST/SLOW) ==> 4.665, 1, 22, .0419
TRAINING (0-90 MIN) ==> 1.558, 3, 66, .207
INTERACTION ==> 18.404, 3, 66, 2.64E-08

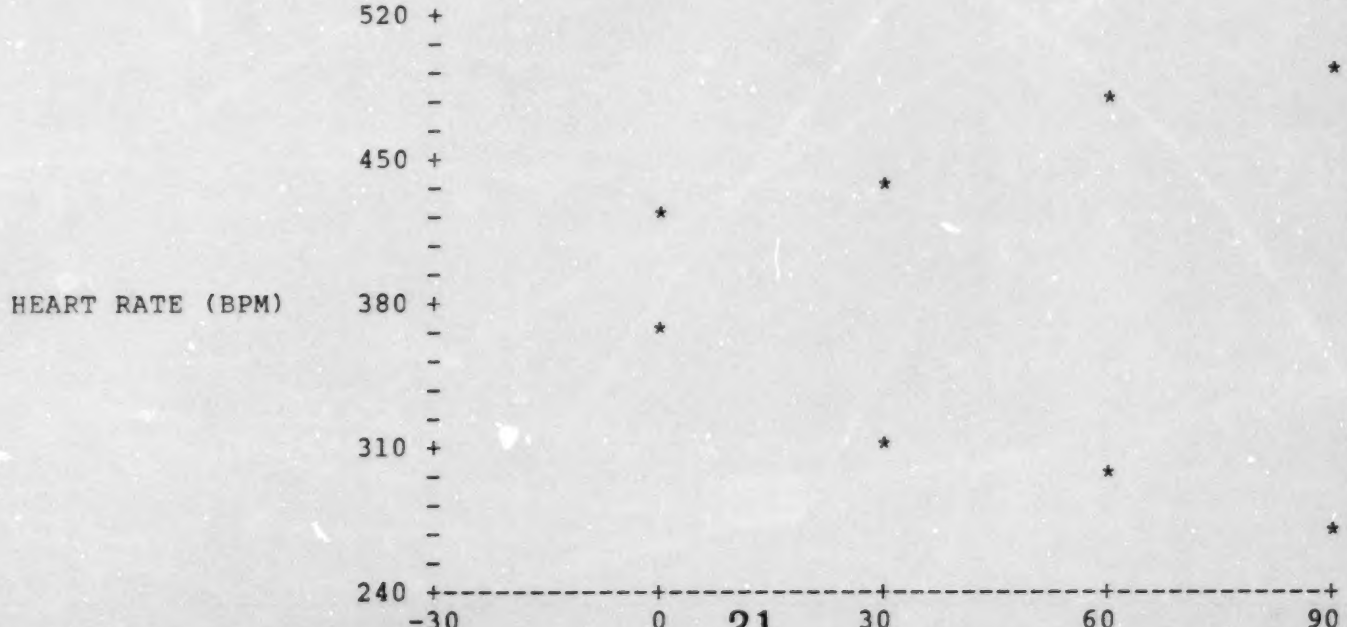
SEED = 2753



MINUTES OF TRAINING

F-RATIO, NUMERATOR DF, DENOMINATOR DF, AND PROBABILITY FOR:
REWARD (FAST/SLOW) ==> 11.271, 1, 22, .00285
TRAINING (0-90 MIN) ==> .485, 3, 66, .698
INTERACTION ==> 9.958, 3, 66, 2.38E-05

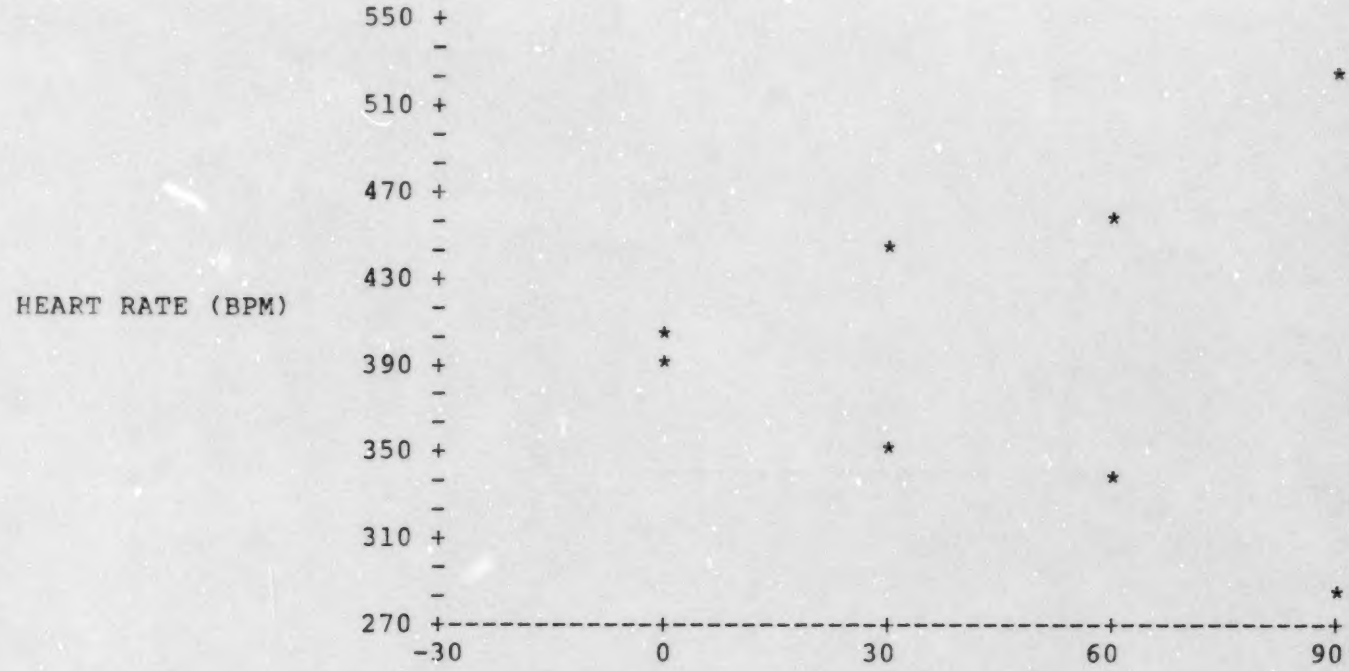
SEED = 1197



MINUTES OF TRAINING

F-RATIO, NUMERATOR DF, DENOMINATOR DF, AND PROBABILITY FOR:
REWARD (FAST/SLOW) ==> 12.28, 1, 22, .002
TRAINING (0-90 MIN) ==> .749, 3, 66, .53
INTERACTION ==> 19.687, 3, 66, 1.09E-08

SEED = 8807



F-RATIO, NUMERATOR DF, DENOMINATOR DF, AND PROBABILITY FOR:
REWARD (FAST/SLOW) ==> 8.602, 1, 22, .0077
TRAINING (0-90 MIN) ==> .216, 3, 66, .885
INTERACTION ==> 20.519, 3, 66, 6.24E-09

? quit